

# SPECIFICATION

Electronic Version 1.2.8

Stylesheet Version 1.0

## **METHOD FOR PROVIDING RELEVANT SEARCH RESULTS BASED ON AN INITIAL ONLINE SEARCH QUERY**

### Background of the Invention

- [0001] The present invention relates generally to the production of Internet search results. More particularly, the present invention relates to generating and retrieving relevant search results based upon an initial search query utilizing a conventional Internet search engine.
- [0002] In its purest form, a search query is no more than a word or phrase. However, such a simple search query usually results in the retrieval of an overabundance of documents, many of which are generally irrelevant but were retrieved nonetheless. In essence, the success and usefulness of a search query depends on the searcher's skill and knowledge in creating and selecting the most accurate words for the search query, as well as the capability of the search engine in providing relevant documents based upon that search query.
- [0003] The amount of informational content available on the Internet is and will, in all probability, continue to expand at an exponential rate. This expansion, coupled with the decentralized and anarchistic nature of the Internet, creates considerable difficulty in locating and retrieving particular informational content.
- [0004] As a result, many existing Internet search providers maintain generalized content based searching. For example, keywords or metatags located in the Internet documents are customarily used wherein the search provider matches the search term with documents containing matching keywords or metatags. However, even when

content is found through an existing Internet search provider, a further difficulty occurs in trying to evaluate the relative merit or relevance of the documents that are retrieved. The search for specific documents utilizing only a few keywords will almost always identify documents whose relevancy is uncertain. Thus, the total volume of irrelevant documents retrieved in the return lists tends to weaken the usefulness of the Internet in finding specific informational content.

[0005] Internet search providers typically seek out and scan the Internet to create objective indexes of Internet sites that can later be searched in response to a searcher's particular query. In order to be recognized as a valuable document locator within the Internet community, the search provider must be capable of performing full searches of all the available information on the Internet, provide immediate search-query response times, and develop an appropriate system for ranking the documents according to their relevancy, amongst other things.

[0006] Once the service provider has indexed individual Internet pages from various Internet sites, the service provider then stores a list of terms, or individual words, that occur or repeat themselves within the indexed pages. In theory, the more frequent certain words appear or repeat within the document, excluding of course simple verbs, prepositions, and conjunctions, the more relevancy those words are given in describing the content of the document. Thus, the greater a certain word appears within an indexed document, the more relevant that document becomes to a searcher who enters that specific word as his or her keyword for a search query.

[0007] However, documents posted on the Internet are often posted with little or no editorial supervision. As a result, many documents are overwhelmed with discrepancies and mistakes that decrease the usefulness of a search engine. In addition, because the Internet has become a medium for advertisers, many Internet sites seek to catch the attention of visitors. As a result, promoters of these sites attempt to incorporate undetectable words, which act as an enticement for drawing the attention of search engines relying on its false relevancy.

[0008] The unreliability associated with many documents on the Internet poses a serious problem when a search engine tries to rank the relevance of located documents. Typically, all that the search engine has to work with is the distribution of words, and

as such, it can do little more than indicate whether or not the distribution of words in a particular document matches the search query more closely than the distribution of words in another document. Furthermore, because there are no standards for relevancy rankings on the Internet, there is no assurance that the highest ranked document returned by a search engine is the most relevant. As such, the uninhibited nature of documents posted on the Internet results in an atmosphere that is not reliably searchable in a well-organized manner by existing search engines.

[0009] Some search engines have attempted to rectify this problem by using a combination of criteria and algorithms to determine the rank and relevance of a particular Internet site for any given search term. For example, some search engines consider the number of links or hyperlinks from a particular Internet site A pointing to another Internet site B as a credit for trustworthiness or importance of site B. Thus, the more links pointing to site B the more relevant that site becomes. These search engines also take into account the importance of site A by analyzing how many links are referring to that specific site. Credits cast by Internet sites that are themselves trustworthy are given more weight in determining the ranking of other sites.

[0010] However, basing a site's ranking or relevancy on the number of links pointing to or from it is subject to the same type of manipulation as other search engine methodologies. For example, Internet site promoters can purchase or participate in link exchange programs wherein they pay another site to refer back to them, thus undermining the very purpose of using links as a form of legitimacy. Furthermore, because these search engines employ an indexing methodology without discriminating against Internet sites that are not heavily trafficked by Internet searchers, their return lists can contain millions of Internet sites, which is impossible for the searcher to comprehend.

[0011] The present invention overcomes the disadvantages and/or shortcomings of known prior art online search engines and provides significant improvements thereover.

## Brief Summary of the Invention

[0012]

It is therefore an object of the present invention to provide a method for reliably

providing relevant search results to a searcher after submitting an initial search query.

- [0013] It is yet another object of the present invention to provide the searcher with the most popular category(ies) for an initial keyword query.
- [0014] A further object of the present invention is to provide the searcher with the most relevant Internet sites based on statistical analysis that tracks the number of times searchers visit a particular Internet site, thus enhancing each Internet site's popularity.
- [0015] Another object of the present invention is to limit the amount of Internet sites that are returned on the return list to a manageable number for the searcher to review.
- [0016] Still a further object of the present invention is to track searcher activities when utilizing the service provider's search engine to determine which Internet sites are visited most within a given category and implement that data into an evolving system that will update the database and provide the searcher's with the most relevant Internet site(s) for any given search term based upon prior results.
- [0017] The present invention is a unique and novel process for conducting Internet based document searches through an Internet search engine by providing a method for reliably and efficiently supplying relevant Internet sites based on an initial keyword query.
- [0018] In an embodiment of the present invention, a searcher enters at least one keyword into a conventional search engine input box. Once the searcher submits the initial search query, the present invention produces a list of relevant Internet sites based upon that initial search term.
- [0019] An embodiment of the present invention maintains at least one database containing predefined categories. Each category contained in this database is created, defined and maintained by a search engine provider or other 3rd party. Once the categories are defined, anticipated search terms provided in a search term database are matched to at least one of the categories based on the search term's definitional relevancy and/or linguistic usage compared to the category. A third database providing Internet site information, such as hyperlink, title, or content, is used wherein each Internet site is matched to at least one of the predefined categories

based upon either an objective or subjective approximated relationship between the content of the documents and/or the predefined descriptions of each respective category.

[0020] Once the searcher submits an initial search query, only the most popular category out of all the categories that the search term may belong to is provided by implementing a set of preponderance criteria that, amongst other means, calculates the number of times a particular category is selected by prior searchers in association with each respective search term used in the initial search, uses subjective determinations made by the search engine provider as to which search terms belong to which categories, and/or calculates the number of times a search term is repeated within the pre-designated keywords contained within Internet sites associated with the category.

[0021] After the searcher submits his or her initial query, a search result list comprising Internet sites belonging to the most popular category is displayed, preferably arranged by relevancy and popularity.

### Brief Description of the Several Views of the Drawings

[0022] The preferred embodiment is herein described in detail with references to the drawings, where appropriate, wherein:

[0023] FIG. 1 is a flowchart depicting the preferred embodiment's method of categorizing Internet sites contained within the Internet site database;

[0024] FIG. 2 is a flowchart depicting an existing typical Internet search result retrieval system;

[0025] FIG. 3 is a flowchart depicting the preferred embodiment's method of selecting the most relevant Internet sites for any given keyword query;

[0026] FIG. 4 is a flowchart depicting the preferred embodiment's method of determining the most relevant category;

[0027] FIG. 5 is a flowchart depicting the preferred embodiment's method of determining the most relevant Internet sites for each category; and

[0028] FIG. 6 is an example of a specific embodiment of the present invention.

## Detailed Description of the Invention

[0029] The preferred embodiment of the present invention is a unique and novel process for conducting Internet based document searches through an Internet search engine by providing a method for reliably and efficiently supplying relevant Internet sites based on an initial keyword query. The present invention utilizes a method of assigning search terms and Internet sites to common, pre-defined topical categories in order to accurately and reliably provide the most relevant Internet information available for any given search query.

[0030] The topical categories are preferably defined with a title and a description, somewhat similar to encyclopedic topics and are provided in a topical category database. Alternately, the topical categories can be defined with other cataloging references, such as a numeric cataloging system, computer cataloging system, and the like.

[0031] The preferred embodiment of the present invention provides an Internet site database with information for at least one Internet site. Preferably, the Internet site database contains information relative to each respective Internet site, such as topic, title, content, author, description, and its uniform resource locator. Referring to FIG. 1, the preferred embodiment of the present invention utilizes a subjective determination to systematically assign each Internet site 1 contained within the Internet site database to at least one pre-defined topical category in the topical category database utilizing a preferred method wherein the Internet site 1 is dissected into 4 subparts; a description 1a, a title 1b, content 1c, and meta-tags 1d. The subparts are used by the search engine service provider to evaluate the Internet site 2 and compare the components of the Internet site to each topical category contained within the topical category database to assign each Internet site to an appropriate topical category(ies) 4. Alternately, the present invention can categorize the Internet site 1 utilizing any combination of the Internet site's 1 description 1a, title 1b, content 1c, or meta-tags 1d. Still alternately, a Internet site can be assigned to a pre-defined topical category by using any sub-part exclusively. In any event, each Internet site 1 is assigned to at least one topical category.

[0034] If more than one pre-defined category is assigned to the initial search term, then the preferred embodiment of the present invention utilizes a preponderant method 10 and determines the most popular topical category for that initial search term.

[0036] Referring to FIGS. 4 and 5, once the preponderant method determines the most

popular topical category, the present invention may utilize statistical market research data to determine the most popular Internet sites assigned to that particular most popular topical category. The service provider may disclose the most popular topical category information with the statistical data provider 19 so that the Internet site information gets assigned to the proper topical category. As such, the present invention may track searcher activities when utilizing the service provider's search engine to determine which Internet sites are visited most within any given topical category and implement that data into an evolving system that will update the topical category database and provide the searchers with the most relevant Internet site(s) for any given search term based upon prior results. Once this statistical information is received by the statistical data provider, the Internet sites may then be organized based on a number of criteria including, but not limited to, the number of unique visitors to each Internet site, the total amount of traffic to the Internet site, the number of hyperlinks pointing to the Internet site, and any other data used to assess the popularity of the Internet site. Once the Internet site(s) are organized for that particular topical category, the most popular topical category is displayed along with its correspondingly assigned Internet site(s) information, followed by the next most popular category with its correspondingly assigned Internet site information in the same fashion as stated above, and so on. Preferably, the most popular topical category search results are listed first by listing all Internet sites' information assigned to that specific most popular topical category and organized with the statistical information. It is anticipated that the most popular topical category assigned to the initial search result will contain the information that the searcher was initially searching for.

[0037]

The present invention can be better illustrated with the following example, which is intended to explain, and not limit, the invention. Referring to FIG. 6, an example embodiment of the present invention reflecting the use of the present invention in an Internet search engine setting, where a searcher enters an initial search term 5 "hotels" in the search input box, submits the search to the search engine provider, and a return list is displayed showing the most popular topical category 9 and its correspondingly assigned Internet site information with statistical popularity data 11. Following the most popular topical category 9 is the second most popular category 22



for the term "hotels" with its correspondingly assigned Internet site information.

[0038] While preferred and alternate embodiments have been described herein, it is to be understood that these descriptions are only illustrative and are thus exemplifications of the present invention and shall not be construed as limiting. It is to be expected that others will contemplate differences, which, while different from the foregoing description, do not depart from the true spirit and scope of the present invention herein described and claimed.